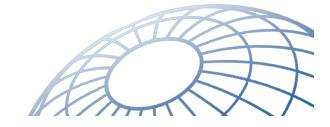
# SnowEx Data Provider Information Session NSIDC DAAC 27 April 2017



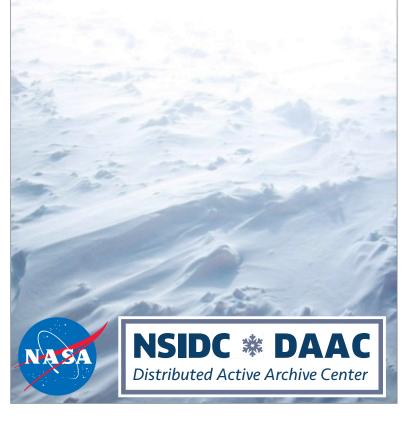
# **Topics**

- Data publication process overview
- Preparing data: Data file requirements & recommendations
- Submitting data to NSIDC
- Publishing data
- Data discovery and access





# NSIDC DAAC Data Management Team



- Amanda Leon Data Management Lead
- Lindsay Sheridan Data Operations
- Amy Fitzgerrell Data & User Support
- Dan Webster Technical Writer
- Mark Schwab Systems Engineer
- Jeff Deems Science Liaison
- Cathy Fowler Database Administrator

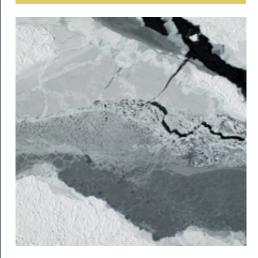
# Data publication process overview

	Data Provider 🛑	DAAC
Prepare data	<ul> <li>Review &amp; implement data file requirements &amp; recommendations</li> <li>Assemble product documentation</li> </ul>	<ul> <li>Assist with data file requirements</li> <li>&amp; recommendations</li> </ul>
Initiate delivery	<ul> <li>Submit data submission form</li> <li>Provide documentation &amp; sample data</li> </ul>	<ul> <li>Review documentation &amp; data file</li> <li>Generate and provide product ID</li> </ul>
Submit data	<ul> <li>Include product ID in filenames</li> <li>Generate MetGen files</li> <li>Transfer data and associated files</li> </ul>	<ul> <li>Assist with MetGen files creation</li> <li>Coordinate data transfer</li> <li>Receive and verify files</li> </ul>
Publish data	Assist with     documentation & data     ingest questions	<ul> <li>Develop user documentation &amp; metadata</li> <li>Run MetGen on data files</li> <li>Ingest data &amp; test distribution</li> </ul>

# Preparing Data: Data File Requirements & Recommendations

- 1. Data formats
- 2. Filenames
- 3. File structure and content

4. General ASCII/CSV structure

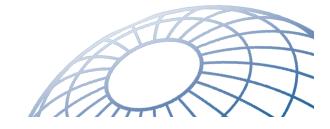


# Required vs. Recommended

While there are few items noted as "required", all recommendations are key to enabling long-term usability of data products

- System requirement Required by NASA data system
- User requirement Essential to data usage





#### NASA Earth Science Approved Formats

https://earthdata.nasa.gov/userresources/standards-and-references#edstandards

**ASCII** HDF5 **HDF-EOS5** NetCDF-4/HDF5 **GeoTIFF** (under review) Distributed Active Archive Center

# 1. File Formats

- Avoid proprietary formats
  - Challenging for users now
  - May not be readable in the future
  - NASA standards allow proprietary for raw/Level-0 data

#### NetCDF4 & HDF5

- Good for multidimensional data
- Capable of holding rich metadata
- Interoperable with variety of computational platforms and protocols (e.g., OPeNDAP)

#### GeoTIFF

- User friendly format; most requested format by NSIDC users
- Widely interoperable with GIS, image processing, and map server applications

# 2. Filenames

- □ Being with data <u>product identifier</u> (NSIDC provides) to capture metrics\*
- ☐ Filenames must be <u>unique</u>\*
- No spaces; ASCII characters only\*
- File <u>extension</u> indicates data format\*
- Include descriptive information such as:
  - Campaign
  - Instrument
  - Measurement
  - Spatial (site, resolution)
  - Temporal (date, time)
  - Processing level or version

<ProductID>\_<ProductSpecificDescriptiveInfo>.<extension>

#### Examples:

ILATM2\_20090422\_131620\_smooth\_nadir3seg\_50pt.csv SMAP\_L3\_FT\_A\_20150418\_R13080\_001.h5













Representation of dates and times



International System of Units



Unit database and conversion between units

**CF Standard Name** 

Climate Forecast (CF) standards promote sharing



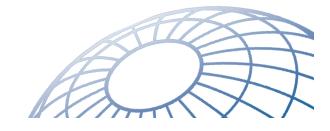
# 3. File structure and content: Variables

- □Define unique, interpretable names for each variable
  - Full interpretation can come from a mapping of short variable names to description
- □Define units for each variable
- Define a value for missing data and use consistently
  - E.g., -9999

# 3. File structure and content: Time\*

- Define time standard and time zone used
  - Recommend using UTC
  - Timestamps may be reported as UTC decimal seconds from the time at which measurements began (commonly as seconds past midnight)
- ☐ Use standard date/time formats
  - Recommend using yyyymmdd, hhmmss, or yyyymmddTHHMMSS.SSSZ (ISO 8601 standard)



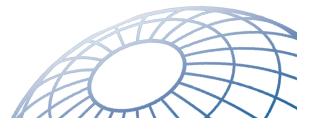


# 3. File structure and content: Geolocation\*

- ☐ Include geographic coordinates for measurements in the data file
- ☐ Use a consistent coordinate format
  - E.g., lat/lon decimal degrees, UTME/UTMN
  - Recommend SnowEx select standard format(s) for delivered data products
- Provide coordinate reference system and horizontal and vertical datum

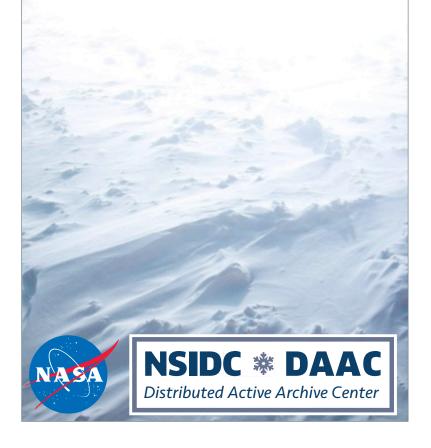






# Full ASCII Earth Science Recommendations

https://earthdata.nasa.gov/sta ndards/ascii-file-formatguidelines-for-earth-sciencedata



# ASCII/CSV

# ASCII/CS\ Structure

- ☐ Include separate header and data sections within file
  - Header needs to be clearly delineated from data rows (e.g., begin with #)
- ☐ Use consistent delimiter between data values
  - Visible characters are preferred (e.g., comma, semi-colon, colon, |)
- □ Separate rows with end-of-line character
  - Mac: CR
  - Unix: LF
  - Windows: CR/LF
- ☐ Do not use empty lines or rows

# **Data File Checklist**

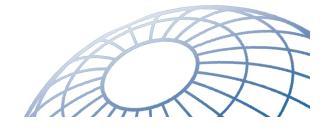
- Data formats
  - ☐ Use nonproprietary formats
  - ☐ Review the NASA Earth Science format recommendations
- Filenames
  - Use descriptive, unique names\*
  - ☐ Product identifier begins each filename\*
  - ☐ No spaces; ASCII characters only\*
  - ☐ File extension indicates data format\*
- File contents/structure: Variables
  - Unique, interpretable variable names
  - Define units for each variable
  - ☐ Consistently use a missing data value

- File contents/structure: Time\*
  - ☐ Define time standard and time zone used: recommend UTC
  - ☐ Use standard date/time formats
- ☐ File structure/content: Geolocation\*
  - ☐ Include geographic coordinates
  - Use a consistent coordinate format
  - ☐ Provide coordinate reference system and horizontal and vertical datum
- ☐ General ASCII/CSV Structure
  - ☐ Include and delineate header from data section
  - Use consistent delimiter between data values: visible characters preferred
  - ☐ Separate rows with EOL
  - Do not use empty lines or rows



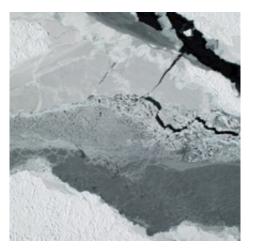
\*System requirement

\*User requirement



# **Submitting Data to NSIDC**

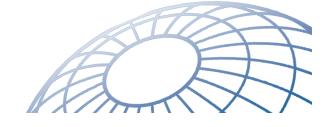
- 1. Data submission form
- 2. Documentation
- 3. MetGen Utility
- 4. Data transfer & ingest



# **Data Submission Form**

- Notifies NSIDC that a data provider is ready to initiate the data publication process
- □ Submitted for each data product a provider intends to deliver
- □ Supplies essential product documentation and a sample data file
- https://nsidc.org/data/snowex/submit

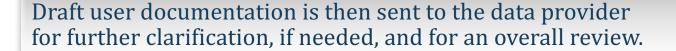




# **Documentation**

#### Dan Webster, Technical Writer

Information provided on the SnowEx Data Submission Form is used by NSIDC to create draft user the documentation, metadata, data citation, and Digital Object Identifier (DOI).



Documentation is then finalized at NSIDC and published in coordination with the publication of the data.

Data set landing page provides access to user documentation, metadata, data citation, and DOI.

• Example: <a href="https://nsidc.org/data/IDCSI4">https://nsidc.org/data/IDCSI4</a>







# **MetGen Utility**

## Mark Schwab, Systems Engineer

#### The MetGen Utility creates two major outputs

# Standards-based ODL or ISO 19115 XML metadata files

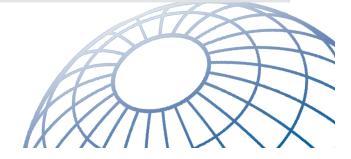
- Required by the NASA data system at NSIDC to ingest data
- Includes a concise geopolygon representation of the data file spatial coverage from any number of longitudelatitude point pairs.

# Product Delivery Record (PDR) files

- Used to automate ingest into the NASA data system
- Used to verify data integrity during transfer and ingest (e.g. checksums)



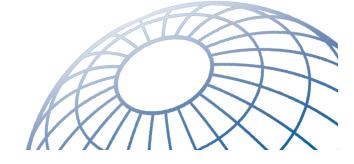




# **MetGen Utility**

- MetGen was developed to be ran by data providers (e.g., IceBridge)
- For SnowEx Year 1, NSIDC will run MetGen
- Need data providers to create and deliver the corresponding MetGen input files for their data files
  - NSIDC will assist providers in creating these





# **MetGen Utility**

# Input files for each data file

PREMET	FILE	EXAMP	LE
1 1 1 1 1 1 1		T-14 PV PT - T T	

VersionID\_local=001

Begin\_date=2014-10-23

End\_date=2014-10-23

Begin\_time=18:11:54.518000

End\_time=18:17:29.073000

Container=AdditionalAttributes

AdditionalAttributeName=ThemeID

ParameterValue=2014\_AN\_NASA

AssociatedPlatformShortName=DC-8

AssociatedInstrumentShortName=ATM

AssociatedSensorShortName=ATM

Data\_FileName=ILATM1B\_20141023\_181210.ATM5BT4.h5

#### SPATIAL FILE EXAMPLE

34.328772 -87.978171

34.329361 -87.978177

34.329945 -87.978184

34.330525 -87.978193

34.331099 -87.978203

34.331668 -87.978214

34.33223 -87.978226

34.332786 -87.978239

34.333337 -87.978253

34.333878 -87.978269

34.334413 -87.978285

34.334941 -87.978302

34.335458 -87.97832

# Data Transfer & Ingest

## Lindsay Sheridan, Data Operations

- Transfer methods:
  - FTP push/pull
  - HTTPS pull
  - Dropbox (compressed folders of files)

- What needs to be delivered:
  - Data files
  - Ancillary files
  - MetGen spatial and premet files
  - Data transfer manifest
- ♦ After receiving the data submission form, Operations will contact providers to assist with the transfer of files.
- For large numbers of files, recommend sending a sample for testing prior to transferring everything.



# Data Transfer & Ingest

#### Data Transfer Manifest Template

```
SnowEx DataSet Manifest
Contact Info: ops@nsidc.org
Instructions: Please fill out the following and send this with the data
that is being provided. Thank you for all of your help.
This Manifest template has been updated as of 4/20/2017 and should be used
for all subsequent data submissions.
_____
Provider Name:
Organization Name:
Contact Email Address:
Date Sent:
Data Set Title:
Data Set ShortName (if known):
Campaign(Region/year):
Is this the first delivery of this Data Set?:
Has the data format changed since the most recent delivery?:
Has the file naming convention changed since the most recent delivery?:
If users will require an updated reader to view these files, is it included in the delivery?:
Method Of Data Transfer (Ftp, Dropbox, etc):
Data Type(s) (.mat, .pdf, .jpeg, etc):
Total Number Of Files:
Total Size Of Data (in MB):
Type Of Checksum (MD5 or CKSUM):
Please include the following for each file that is being sent:
FileName, FileSize, File Checksum
```





# **Publishing Data**

- Will publish data products as files and documentation are received
- Will accommodate SnowEx priorities
- Publication rate is constrained by team capacity
  - Data surveys indicated:
    - ~25 data providers
    - ~50-60 data products
- Will provide publication status to the SnowEx leadership, providers, and community





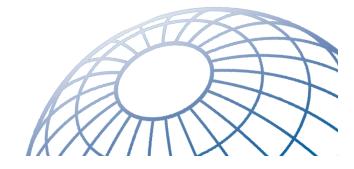
# **Data Discovery & Access**

Amy Fitzgerrell, Data & User Support

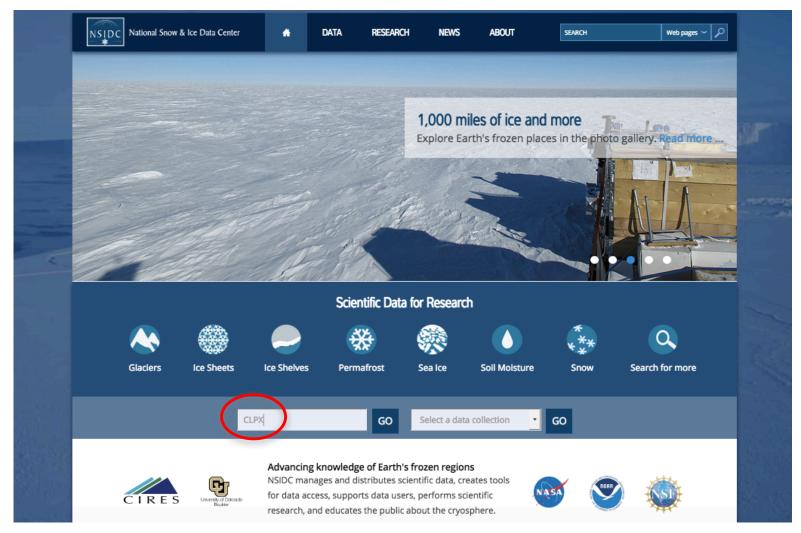
The NSIDC DAAC data system utilizes metadata to enable data discovery.

- NSIDC Search
- NASA Earthdata Search





## **NSIDC Data Search: CLPX**

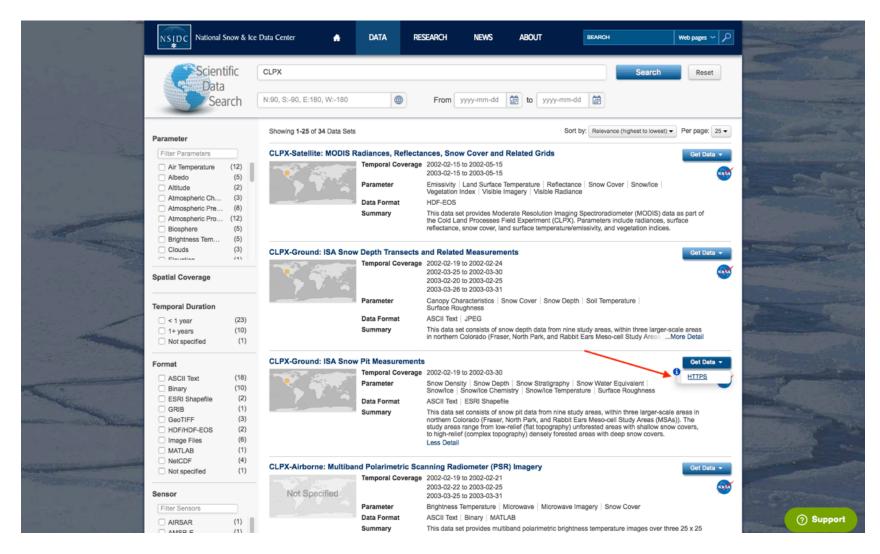


■ We'll perform an NSIDC data search for CLPX, which contains no file-level metadata





## **NSIDC Data Search: CLPX**

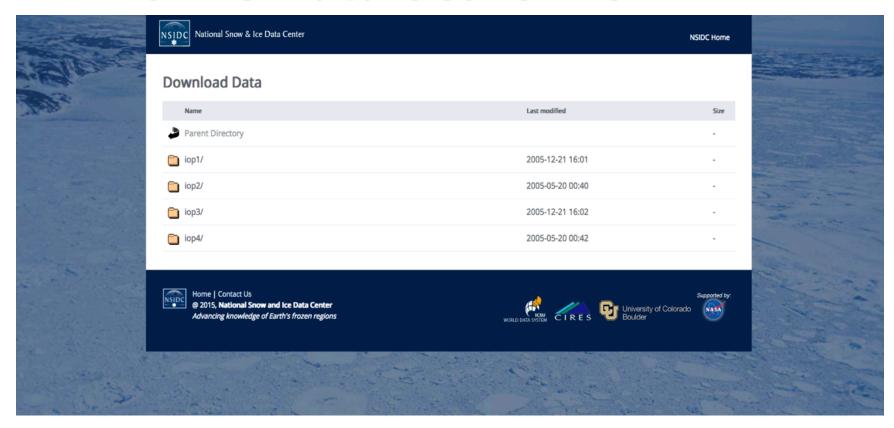


☐ The only "Get Data" option available is HTTPS





#### **NSIDC** Data Search: CLPX

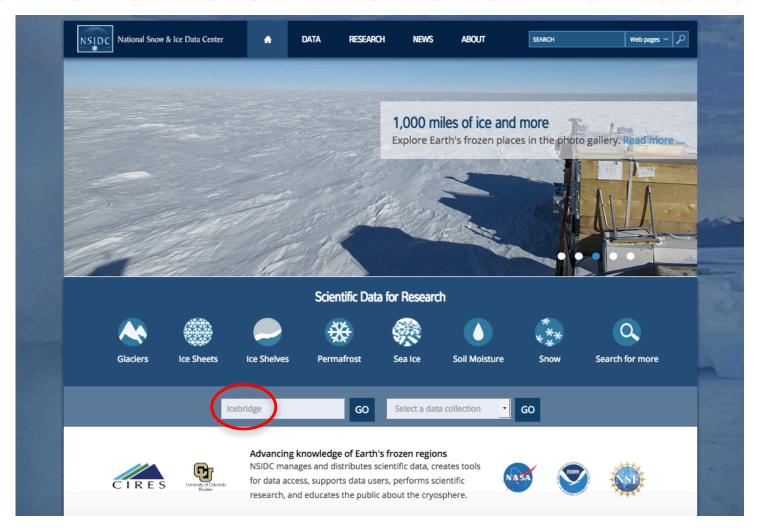


- ☐ This lands a user in the data set's HTTPS directory
- Users explore the directory structure to locate individual data files



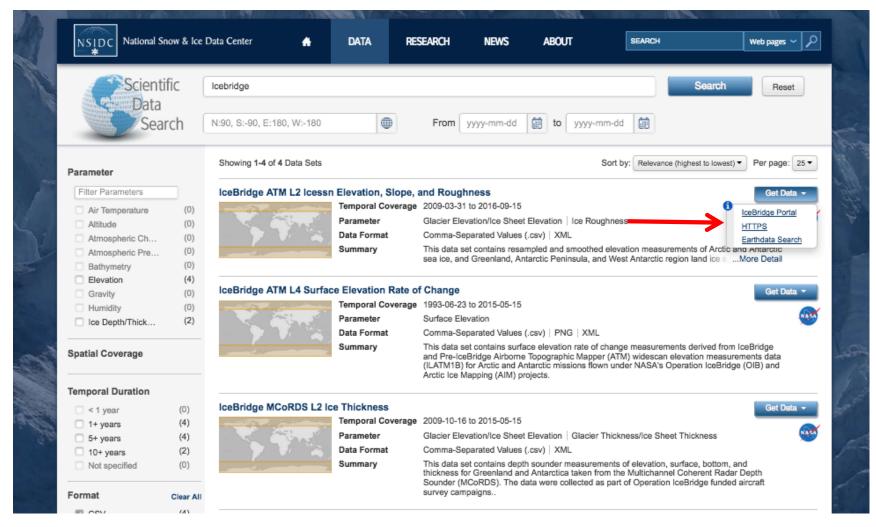






☐ MetGen has been used with many IceBridge data sets to generate file-level metadata



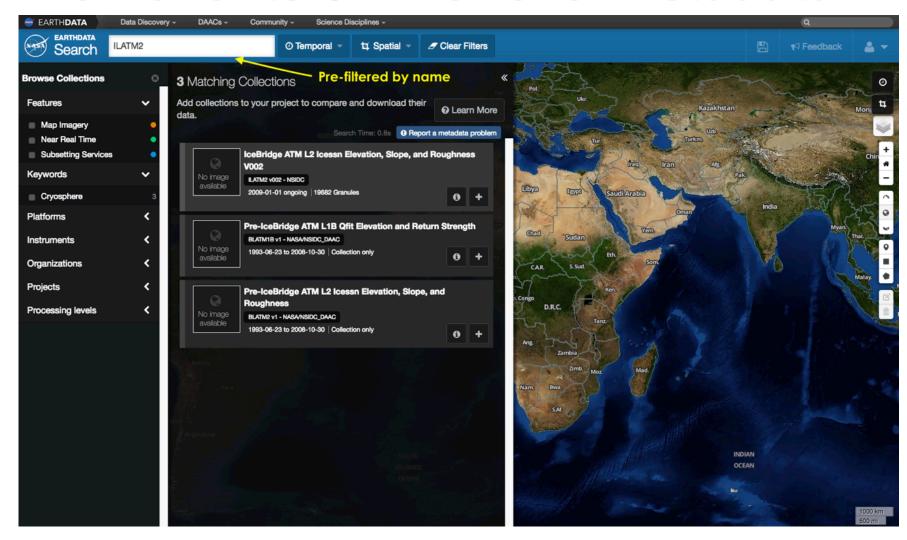


■ More "Get Data" options become available – we'll choose Earthdata Search



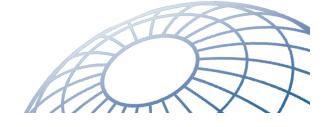


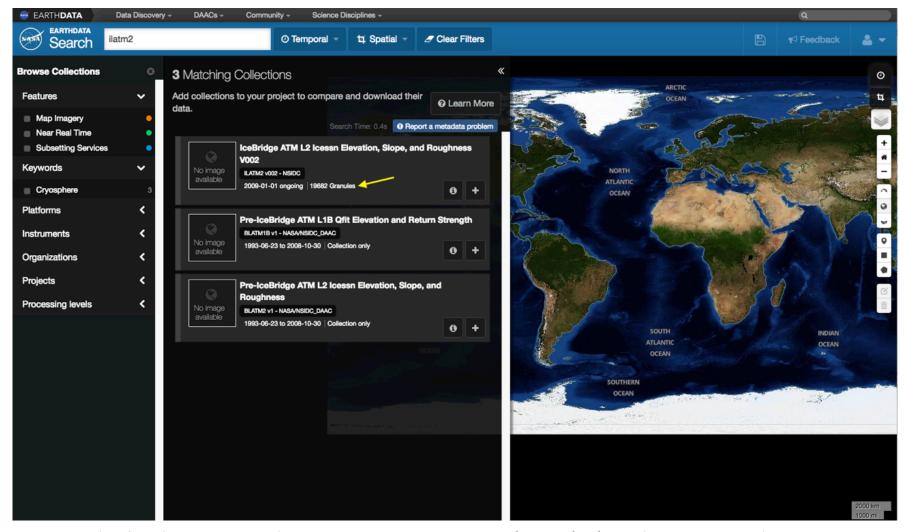




■ Earthdata Search opens, all set to explore ILATM2



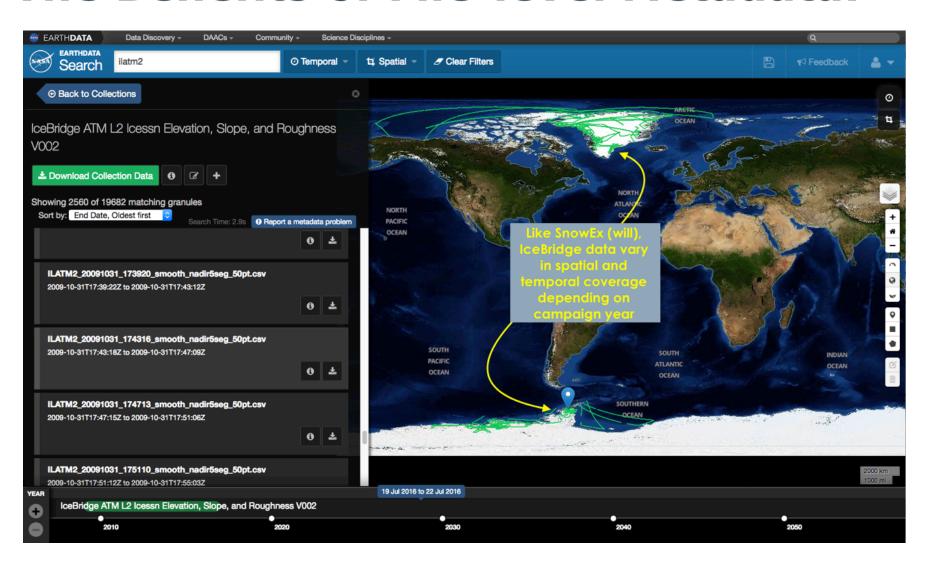




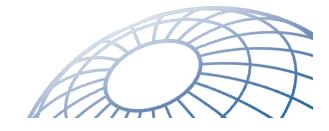
 Click the granule count to expand and display granules in the map

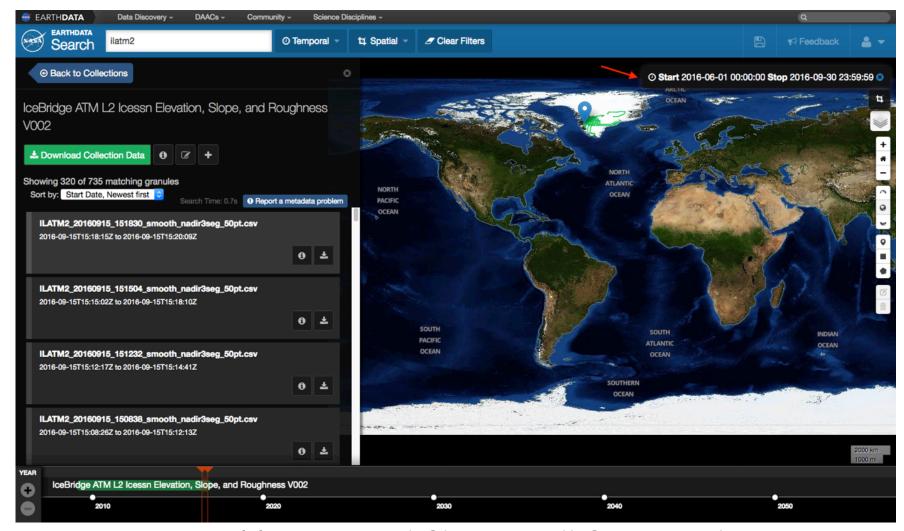








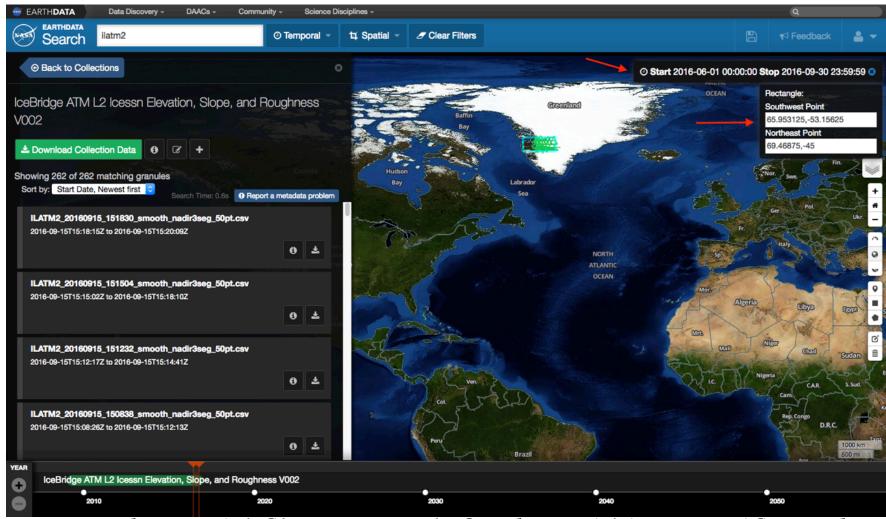




Users can add a temporal filter to cull for, or within, a specific campaign



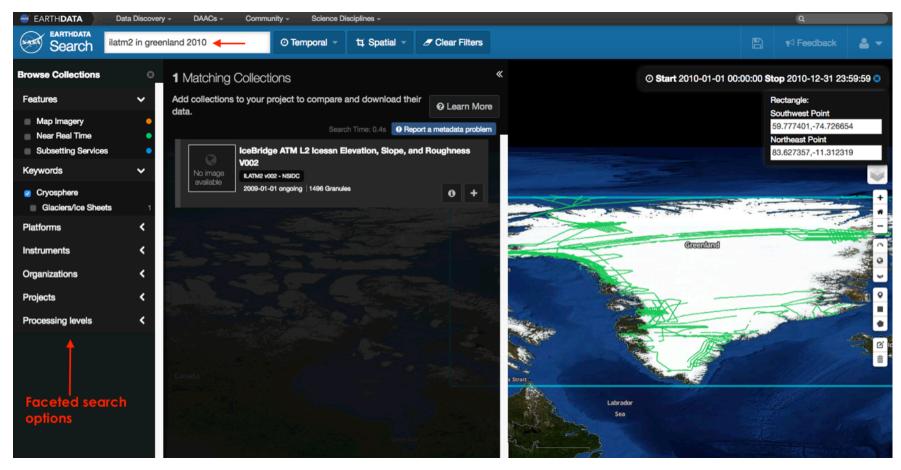




...and a spatial filter to search for data within a specific study area, building an order of just the desired data granules



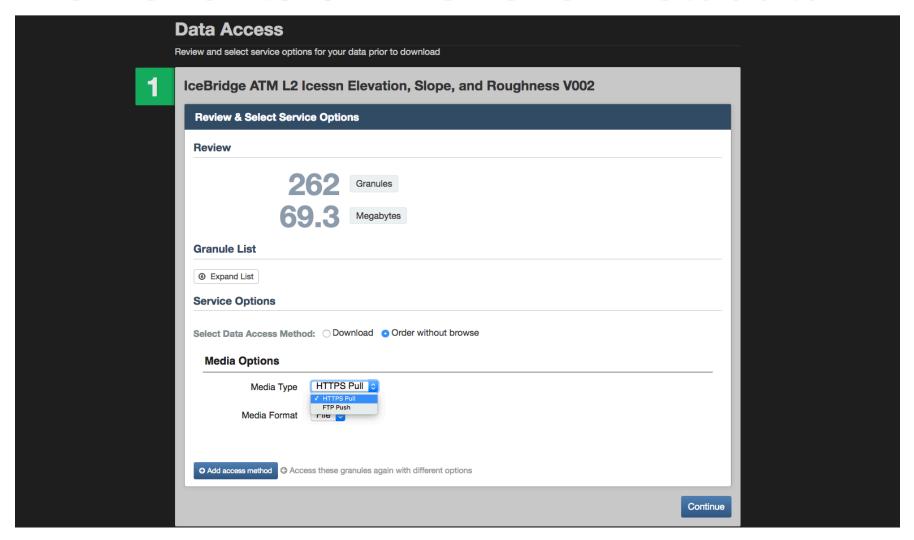




- There's also a "natural language" search option that will add keyword, temporal and spatial filters based on the phrase typed.
- □ Faceted search options allow the user to filter based on characteristics like instrument or processing level.







Regarding data access, there's a greater choice of order options: Download, HTTPS Pull, or FTP Push





# Direct data access also available

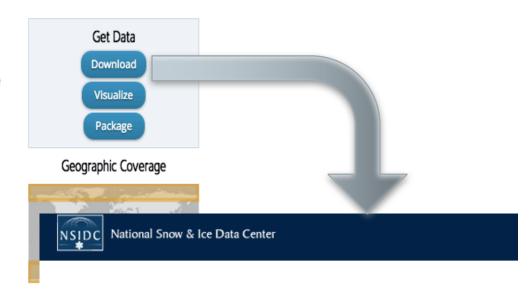
Name



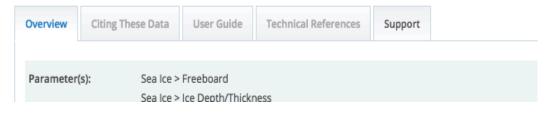
Data Set ID: IDCSI4

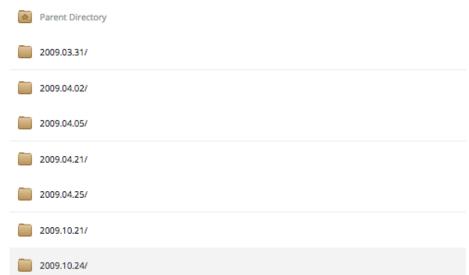
#### IceBridge L4 Sea Ice Freeboard, Snow Depth, and Thickness, Version 1

This data set contains derived geophysical data products including sea ice freeboard, snow depth, and sea ice thickness measurements in Greenland and Antarctica retrieved from IceBridge Snow Radar, Digital Mapping System (DMS), Continuous Airborne Mapping By Optical Translator (CAMBOT), and Airborne Topographic Mapper (ATM) data sets. The data were collected as part of Operation IceBridge funded campaigns.













#### Resources

# NSIDC SnowEx Website: Information for SnowEx Data Providers

https://nsidc.org/data/snowex/information-snowex-data-providers

- Data file requirements & recommendations
- Sample ASCII/CSV data files
- Links to resources
- Data submission form
- Sample MetGen premet and spatial files
- Data manifest template

Email address: snowex@nsidc.org



# **Questions?**

#### **DAAC Data Provider** Review & implement data file Assist with data file **Prepare** requirements & requirements & data recommendations recommendations Assemble product documentation Review documentation & data **Submit data Initiate** submission form delivery Generate and provide product Provide documentation & sample data Include product ID in filenames Assist with MetGen files Generate MetGen files creation Submit Transfer data and associated files Coordinate data transfer data Receive and verify files Develop user documentation & Assist with metadata documentation & data **Publish** Run MetGen ingest questions data Ingest data & test distribution Provide publication status

